



# wwPDB X-ray Structure Validation Summary Report ⓘ

May 22, 2020 – 03:20 am BST

PDB ID : 5J5V  
Title : CdiA-CT from uropathogenic Escherichia coli in complex with cognate immunity protein and CysK  
Authors : Goulding, C.W.; Johnson, P.M.; Morse, R.P.  
Deposited on : 2016-04-04  
Resolution : 2.75 Å(reported)

This is a wwPDB X-ray Structure Validation Summary Report for a publicly released PDB entry.

We welcome your comments at [validation@mail.wwpdb.org](mailto:validation@mail.wwpdb.org)

A user guide is available at

<https://www.wwpdb.org/validation/2017/XrayValidationReportHelp>

with specific help available everywhere you see the ⓘ symbol.

---

The following versions of software and data (see [references ⓘ](#)) were used in the production of this report:

MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.11
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.11

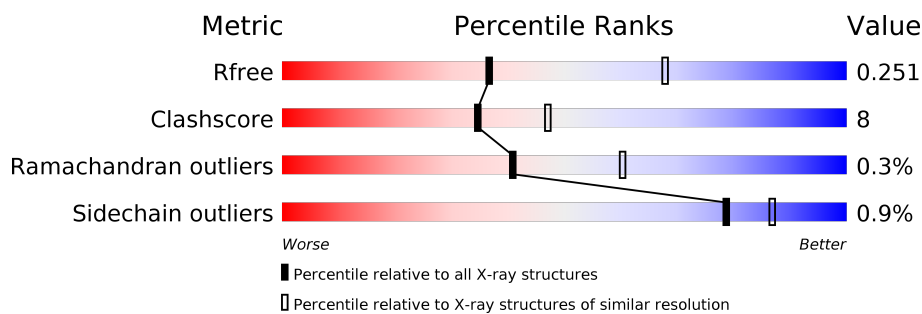
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 2.75 Å.

Percentile scores (ranging between 0-100) for global validation metrics of the entry are shown in the following graphic. The table shows the number of entries on which the scores are based.



Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
$R_{free}$	130704	1235 (2.78-2.74)
Clashscore	141614	1277 (2.78-2.74)
Ramachandran outliers	138981	1257 (2.78-2.74)
Sidechain outliers	138945	1257 (2.78-2.74)

The table below summarises the geometric issues observed across the polymeric chains and their fit to the electron density. The red, orange, yellow and green segments on the lower bar indicate the fraction of residues that contain outliers for  $\geq 3$ , 2, 1 and 0 types of geometric quality criteria respectively. A grey segment represents the fraction of residues that are not modelled. The numeric value for each fraction is indicated below the corresponding segment, with a dot representing fractions  $\leq 5\%$ .

Mol	Chain	Length	Quality of chain
1	A	323	88% 9% .
1	D	323	86% 11% .
2	B	228	29% 13% 58%
2	E	228	36% 6% . 58%
3	C	138	59% 29% . 11%
3	F	138	65% 23% . 9%

## 2 Entry composition

There are 4 unique types of molecules in this entry. The entry contains 8296 atoms, of which 0 are hydrogens and 0 are deuteriums.

In the tables below, the ZeroOcc column contains the number of atoms modelled with zero occupancy, the AltConf column contains the number of residues with at least one atom in alternate conformation and the Trace column contains the number of residues modelled with at most 2 atoms.

- Molecule 1 is a protein called Cysteine synthase A.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
1	A	313	Total	C	N	O	P	S	0	0	0
			2346	1480	404	455	1	6			
1	D	313	Total	C	N	O	P	S	0	0	0
			2346	1480	404	455	1	6			

There are 2 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	2	GLY	SER	engineered mutation	UNP P0ABK6
D	2	GLY	SER	engineered mutation	UNP P0ABK6

- Molecule 2 is a protein called tRNA nuclease CdiA.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	B	96	Total	C	N	O	Se	0	0	0
			750	461	138	148	3			
2	E	96	Total	C	N	O	Se	0	0	0
			750	461	138	148	3			

There are 2 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
B	0	MSE	-	initiating methionine	UNP Q0T963
E	0	MSE	-	initiating methionine	UNP Q0T963

- Molecule 3 is a protein called Immunity protein CdiI.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
3	C	123	Total	C	N	O	S	Se	0	0	0
			987	625	164	194	2	2			
3	F	126	Total	C	N	O	S	Se	0	0	0
			1016	645	169	198	2	2			

There are 22 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
C	1	MSE	-	initiating methionine	UNP Q0T964
C	129	THR	-	expression tag	UNP Q0T964
C	130	SER	-	expression tag	UNP Q0T964
C	131	LEU	-	expression tag	UNP Q0T964
C	132	GLU	-	expression tag	UNP Q0T964
C	133	HIS	-	expression tag	UNP Q0T964
C	134	HIS	-	expression tag	UNP Q0T964
C	135	HIS	-	expression tag	UNP Q0T964
C	136	HIS	-	expression tag	UNP Q0T964
C	137	HIS	-	expression tag	UNP Q0T964
C	138	HIS	-	expression tag	UNP Q0T964
F	1	MSE	-	initiating methionine	UNP Q0T964
F	129	THR	-	expression tag	UNP Q0T964
F	130	SER	-	expression tag	UNP Q0T964
F	131	LEU	-	expression tag	UNP Q0T964
F	132	GLU	-	expression tag	UNP Q0T964
F	133	HIS	-	expression tag	UNP Q0T964
F	134	HIS	-	expression tag	UNP Q0T964
F	135	HIS	-	expression tag	UNP Q0T964
F	136	HIS	-	expression tag	UNP Q0T964
F	137	HIS	-	expression tag	UNP Q0T964
F	138	HIS	-	expression tag	UNP Q0T964

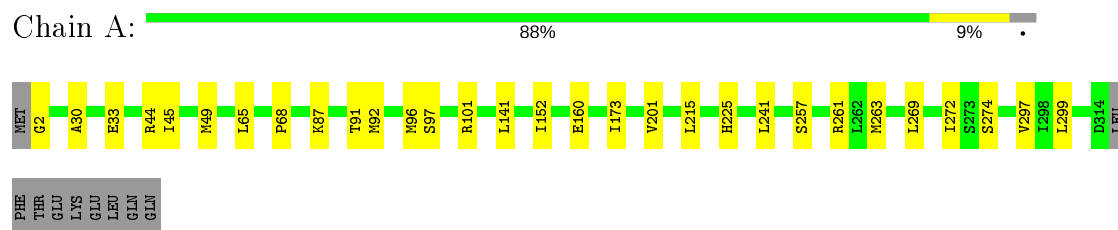
- Molecule 4 is water.

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
4	A	37	Total O 37 37	0	0
4	B	6	Total O 6 6	0	0
4	C	7	Total O 7 7	0	0
4	D	31	Total O 31 31	0	0
4	E	8	Total O 8 8	0	0
4	F	12	Total O 12 12	0	0

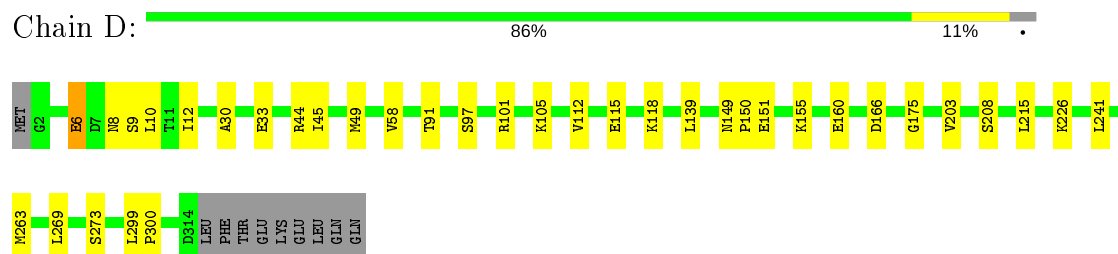
### 3 Residue-property plots [i](#)

These plots are drawn for all protein, RNA and DNA chains in the entry. The first graphic for a chain summarises the proportions of the various outlier classes displayed in the second graphic. The second graphic shows the sequence view annotated by issues in geometry. Residues are color-coded according to the number of geometric quality criteria for which they contain at least one outlier: green = 0, yellow = 1, orange = 2 and red = 3 or more. Stretches of 2 or more consecutive residues without any outlier are shown as a green connector. Residues present in the sample, but not in the model, are shown in grey.

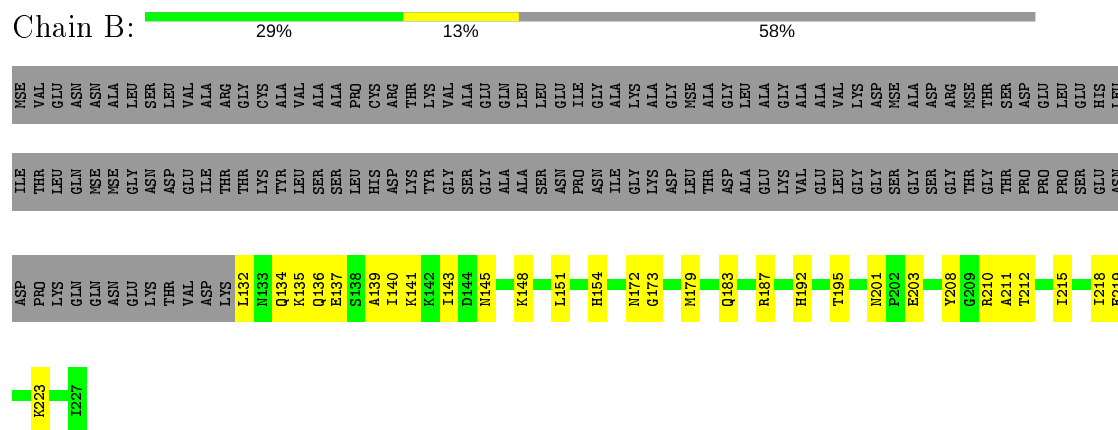
- Molecule 1: Cysteine synthase A



- Molecule 1: Cysteine synthase A

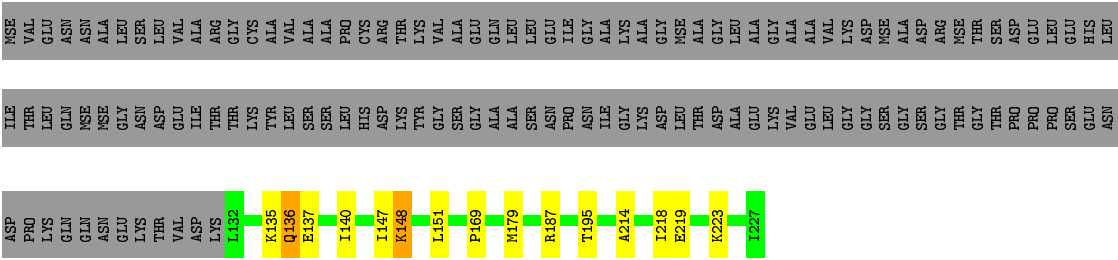


- Molecule 2: tRNA nuclease CdiA



- Molecule 2: tRNA nuclease CdiA

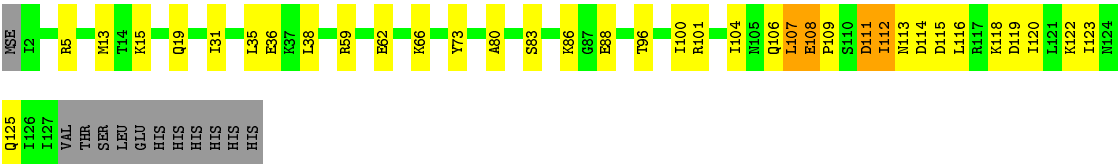




• Molecule 3: Immunity protein CdiI



• Molecule 3: Immunity protein CdiI



## 4 Data and refinement statistics

Property	Value	Source
Space group	C 2 2 21	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	81.25Å 195.54Å 175.06Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	48.83 – 2.75 48.82 – 2.09	Depositor EDS
% Data completeness (in resolution range)	100.0 (48.83-2.75) 76.8 (48.82-2.09)	Depositor EDS
$R_{merge}$	(Not available)	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	0.06 (at 2.08Å)	Xtriage
Refinement program	PHENIX (1.10.1_2155: ???)	Depositor
R, $R_{free}$	0.196 , 0.248 0.201 , 0.251	Depositor DCC
$R_{free}$ test set	2000 reflections (2.49%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	28.8	Xtriage
Anisotropy	0.185	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.30 , 20.4	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.48$ , $\langle L^2 \rangle = 0.32$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
$F_o, F_c$ correlation	0.93	EDS
Total number of atoms	8296	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	34.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.03% of the height of the origin peak. No significant pseudotranslation is detected.*

<sup>1</sup>Intensities estimated from amplitudes.

<sup>2</sup>Theoretical values of  $\langle |L| \rangle$ ,  $\langle L^2 \rangle$  for acentric reflections are 0.5, 0.333 respectively for untwinned datasets, and 0.375, 0.2 for perfectly twinned datasets.

## 5 Model quality [i](#)

### 5.1 Standard geometry [i](#)

Bond lengths and bond angles in the following residue types are not validated in this section: LLP

The Z score for a bond length (or angle) is the number of standard deviations the observed value is removed from the expected value. A bond length (or angle) with  $|Z| > 5$  is considered an outlier worth inspection. RMSZ is the root-mean-square of all Z scores of the bond lengths (or angles).

Mol	Chain	Bond lengths		Bond angles	
		RMSZ	$\# Z  > 5$	RMSZ	$\# Z  > 5$
1	A	0.48	0/2354	0.59	0/3185
1	D	0.50	1/2354 (0.0%)	0.60	0/3185
2	B	0.46	0/757	0.60	0/1011
2	E	0.48	0/757	0.68	0/1011
3	C	0.45	1/996 (0.1%)	0.62	0/1345
3	F	0.50	0/1025	0.77	4/1383 (0.3%)
All	All	0.48	2/8243 (0.0%)	0.63	4/11120 (0.0%)

Chiral center outliers are detected by calculating the chiral volume of a chiral center and verifying if the center is modelled as a planar moiety or with the opposite hand. A planarity outlier is detected by checking planarity of atoms in a peptide group, atoms in a mainchain group or atoms of a sidechain that are expected to be planar.

Mol	Chain	#Chirality outliers	#Planarity outliers
2	B	0	1
2	E	0	1
All	All	0	2

All (2) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	D	208	SER	C-N	5.51	1.44	1.34
3	C	44	CYS	CB-SG	-5.43	1.73	1.81

All (4) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
3	F	111	ASP	N-CA-C	9.40	136.39	111.00
3	F	112	ILE	N-CA-C	-7.53	90.67	111.00
3	F	111	ASP	CB-CA-C	-7.07	96.27	110.40

*Continued on next page...*



*Continued from previous page...*

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
3	F	107	LEU	CA-CB-CG	5.43	127.80	115.30

There are no chirality outliers.

All (2) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
2	B	148	LYS	Peptide
2	E	148	LYS	Peptide

## 5.2 Too-close contacts ⓘ

In the following table, the Non-H and H(model) columns list the number of non-hydrogen atoms and hydrogen atoms in the chain respectively. The H(added) column lists the number of hydrogen atoms added and optimized by MolProbity. The Clashes column lists the number of clashes within the asymmetric unit, whereas Symm-Clashes lists symmetry related clashes.

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
1	A	2346	0	2413	20	0
1	D	2346	0	2413	28	0
2	B	750	0	746	25	2
2	E	750	0	746	17	0
3	C	987	0	1013	34	0
3	F	1016	0	1054	28	0
4	A	37	0	0	0	0
4	B	6	0	0	0	0
4	C	7	0	0	2	0
4	D	31	0	0	2	0
4	E	8	0	0	0	0
4	F	12	0	0	0	0
All	All	8296	0	8385	137	2

The all-atom clashscore is defined as the number of clashes found per 1000 atoms (including hydrogen atoms). The all-atom clashscore for this structure is 8.

The worst 5 of 137 close contacts within the same asymmetric unit are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
2:E:187:ARG:HH12	3:F:116:LEU:HD11	1.27	1.00
2:B:145:ASN:HD21	2:B:210:ARG:HH11	1.26	0.83

*Continued on next page...*

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:D:151:GLU:OE2	1:D:155:LYS:NZ	2.13	0.81
2:B:187:ARG:HH12	3:C:116:LEU:HD11	1.46	0.80
3:F:101:ARG:NH2	3:F:123:ILE:O	2.15	0.78

All (2) symmetry-related close contacts are listed below. The label for Atom-2 includes the symmetry operator and encoded unit-cell translations to be applied.

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
2:B:154:HIS:CE1	2:B:154:HIS:CE1[3_654]	1.07	1.13
2:B:154:HIS:ND1	2:B:154:HIS:CE1[3_654]	1.63	0.57

## 5.3 Torsion angles [i](#)

### 5.3.1 Protein backbone [i](#)

In the following table, the Percentiles column shows the percent Ramachandran outliers of the chain as a percentile score with respect to all X-ray entries followed by that with respect to entries of similar resolution.

The Analysed column shows the number of residues for which the backbone conformation was analysed, and the total number of residues.

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
1	A	310/323 (96%)	300 (97%)	10 (3%)	0	100	100
1	D	310/323 (96%)	300 (97%)	10 (3%)	0	100	100
2	B	94/228 (41%)	89 (95%)	5 (5%)	0	100	100
2	E	94/228 (41%)	89 (95%)	5 (5%)	0	100	100
3	C	121/138 (88%)	107 (88%)	12 (10%)	2 (2%)	9	16
3	F	124/138 (90%)	113 (91%)	10 (8%)	1 (1%)	19	34
All	All	1053/1378 (76%)	998 (95%)	52 (5%)	3 (0%)	41	60

All (3) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
3	C	111	ASP
3	F	108	GLU
3	C	108	GLU

### 5.3.2 Protein sidechains ⓘ

In the following table, the Percentiles column shows the percent sidechain outliers of the chain as a percentile score with respect to all X-ray entries followed by that with respect to entries of similar resolution.

The Analysed column shows the number of residues for which the sidechain conformation was analysed, and the total number of residues.

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles	
1	A	246/257 (96%)	245 (100%)	1 (0%)	91	93
1	D	246/257 (96%)	245 (100%)	1 (0%)	91	93
2	B	79/172 (46%)	78 (99%)	1 (1%)	69	81
2	E	79/172 (46%)	78 (99%)	1 (1%)	69	81
3	C	113/127 (89%)	112 (99%)	1 (1%)	78	87
3	F	117/127 (92%)	114 (97%)	3 (3%)	46	66
All	All	880/1112 (79%)	872 (99%)	8 (1%)	78	87

5 of 8 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	D	6	GLU
3	F	125	GLN
3	F	114	ASP
3	C	83	SER
2	E	136	GLN

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (2) such sidechains are listed below:

Mol	Chain	Res	Type
3	C	99	GLN
2	E	136	GLN

### 5.3.3 RNA ⓘ

There are no RNA molecules in this entry.

## 5.4 Non-standard residues in protein, DNA, RNA chains ⓘ

2 non-standard protein/DNA/RNA residues are modelled in this entry.

In the following table, the Counts columns list the number of bonds (or angles) for which Mogul statistics could be retrieved, the number of bonds (or angles) that are observed in the model and the number of bonds (or angles) that are defined in the Chemical Component Dictionary. The Link column lists molecule types, if any, to which the group is linked. The Z score for a bond length (or angle) is the number of standard deviations the observed value is removed from the expected value. A bond length (or angle) with  $|Z| > 2$  is considered an outlier worth inspection. RMSZ is the root-mean-square of all Z scores of the bond lengths (or angles).

Mol	Type	Chain	Res	Link	Bond lengths			Bond angles		
					Counts	RMSZ	# $ Z  > 2$	Counts	RMSZ	# $ Z  > 2$
1	LLP	D	42	1	23,24,25	2.34	5 (21%)	25,32,34	1.61	5 (20%)
1	LLP	A	42	1	23,24,25	2.23	5 (21%)	25,32,34	1.60	4 (16%)

In the following table, the Chirals column lists the number of chiral outliers, the number of chiral centers analysed, the number of these observed in the model and the number defined in the Chemical Component Dictionary. Similar counts are reported in the Torsion and Rings columns. '-' means no outliers of that kind were identified.

Mol	Type	Chain	Res	Link	Chirals	Torsions	Rings
1	LLP	D	42	1	-	6/16/17/19	0/1/1/1
1	LLP	A	42	1	-	3/16/17/19	0/1/1/1

The worst 5 of 10 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	D	42	LLP	C4'-NZ	6.80	1.50	1.27
1	A	42	LLP	C4'-NZ	6.42	1.48	1.27
1	D	42	LLP	C4-C4'	5.82	1.57	1.46
1	A	42	LLP	C4-C4'	5.65	1.57	1.46
1	D	42	LLP	O3-C3	3.65	1.45	1.37

The worst 5 of 9 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	D	42	LLP	OP4-C5'-C5	4.36	117.65	109.35
1	D	42	LLP	CE-NZ-C4'	-4.10	106.30	118.90
1	A	42	LLP	CE-NZ-C4'	-3.80	107.23	118.90
1	A	42	LLP	C5-C4-C4'	-2.71	117.09	121.56
1	A	42	LLP	CD-CE-NZ	-2.42	105.00	110.93

There are no chirality outliers.

5 of 9 torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
1	D	42	LLP	CD-CE-NZ-C4'
1	D	42	LLP	C3-C4-C4'-NZ
1	A	42	LLP	C4-C4'-NZ-CE
1	D	42	LLP	CA-CB-CG-CD
1	A	42	LLP	CG-CD-CE-NZ

There are no ring outliers.

No monomer is involved in short contacts.

## 5.5 Carbohydrates [i](#)

There are no carbohydrates in this entry.

## 5.6 Ligand geometry [i](#)

There are no ligands in this entry.

## 5.7 Other polymers [i](#)

There are no such residues in this entry.

## 5.8 Polymer linkage issues [i](#)

There are no chain breaks in this entry.

## 6 Fit of model and data ⓘ

### 6.1 Protein, DNA and RNA chains ⓘ

Unable to reproduce the depositors R factor - this section is therefore empty.

### 6.2 Non-standard residues in protein, DNA, RNA chains ⓘ

Unable to reproduce the depositors R factor - this section is therefore empty.

### 6.3 Carbohydrates ⓘ

Unable to reproduce the depositors R factor - this section is therefore empty.

### 6.4 Ligands ⓘ

Unable to reproduce the depositors R factor - this section is therefore empty.

### 6.5 Other polymers ⓘ

Unable to reproduce the depositors R factor - this section is therefore empty.